

Pengembangan Perangkat Lunak Prediktor Nilai Mahasiswa Menggunakan Metode Spectral Clustering dan Bagging Regresi Linier

Ahmad Yusuf, Hari Ginardi dan Isye Ariesanti

Teknik Informatika, Fakultas Teknologi Informasi, Institut Teknologi Sepuluh Nopember (ITS)

Jl. Arief Rahman Hakim, Surabaya 60111

E-mail: hari@its-sby.edu

Abstrak— Salah satu aplikasi penggalian data dalam dunia pendidikan adalah prediksi nilai ujian akhir mahasiswa. Prediksi nilai ujian akhir dapat membantu pendidik atau mahasiswa dalam melakukan tindakan ke depan yang sesuai dengan kondisi pada saat itu. Salah satu metode yang dapat digunakan dalam memprediksi nilai adalah dengan Regresi Linier. Tetapi, persamaan regresi terkadang dipengaruhi oleh keragaman persebaran data, sehingga keakuratan persamaan prediksi yang dihasilkan kurang baik. Oleh karena itu, ada studi lain yang mengusulkan model Bootstrap Aggregating Regresi Linear yaitu metode regresi linier yang didahului dengan mengelompokkan data. Dalam pengelompokan data digunakan Spectral Clustering. Dengan adanya pengelompokan tersebut, maka diharapkan distribusi data dalam membentuk persamaan regresi dapat lebih seragam. Selanjutnya, setiap kelompok data akan menghasilkan suatu persamaan regresi. Hasil prediksi merupakan rata-rata dari hasil persamaan regresi dari masing-masing kelompok data. Dalam studi ini, dikembangkan sebuah perangkat lunak prediktor nilai mahasiswa menggunakan model yang diusulkan oleh studi lain yang telah disebutkan sebelumnya. Berdasar pada uji coba yang dilakukan, perangkat lunak yang dikembangkan dengan menggunakan Spectral Clustering dan Bootstrap Aggregating Regresi Linier mampu memprediksi nilai akhir mahasiswa dengan parameter jumlah cluster yang tepat. Hal ini dapat disimpulkan berdasarkan nilai kesalahan dengan Root Mean Square Error dari hasil prediksi sekitar 0.05 – 0.08 dari dataset yang digunakan.

Kata Kunci— Bootstrap Aggregating, Data Mining, Prediksi Nilai, Regresi Linear, Spectral Clustering

I. PENDAHULUAN

Penggalian data di dunia pendidikan (*educational data mining*) merupakan hal penting karena penyelenggara pendidikan bisa mendapatkan informasi terkait dari data yang berjumlah besar dan beragam. Educational Data Mining (EDM) bertujuan untuk membangun model komputasional dalam penggalian informasi yang berkaitan dengan pendidikan. Dalam EDM diperlukan metode yang baik dalam memperoleh informasi yang tersimpan.

Salah satu aplikasi penggalian data dalam dunia pendidikan adalah prediksi nilai mahasiswa. Nilai merupakan parameter kesuksesan seorang mahasiswa. Oleh karena itu, prediksi nilai ini sangat penting dalam membantu penyelenggara pendidikan dalam melakukan tindakan ke

depan yang sesuai.

Model yang umumnya digunakan dalam prediksi adalah Regresi Linier. Dengan regresi akan didapatkan nilai prediksi sesuai dengan variabel-variabel yang berpengaruh. Hanya saja *Noise* yang terlalu banyak pada data dapat menyebabkan model regresi yang dibentuk kurang akurat. Hal ini menyebabkan prediksi yang dihasilkan kurang baik.

Akurasi regresi dapat diperbaiki dengan mengatasi *noise* pada data. *Noise* dapat ditangani dengan metode Clustering (pengelompokan) data. Dengan Clustering, data akan dikelompokkan berdasarkan kesamaan data sehingga *noise* pada data latih dapat diminimalisasi. Harapannya akurasi dari model regresi yang dihasilkan akan lebih baik.

Algoritma Clustering yang ada pun bermacam-macam, sebagai contoh K-Means Clustering, Fuzzy C-Means Clustering, dan sebagainya. Algoritma clustering yang sangat umum digunakan adalah K-Means Clustering. Metode K-means mudah dalam pengimplementasiannya serta memiliki waktu komputasi yang cukup cepat. Tetapi metode ini mempunyai kelemahan dalam menganalisis persebaran data serta bergantung pada inisialisasi centroid. K-means hanya melihat jarak data ke masing-masing centroid pada setiap cluster. Salah satu metode clustering lain yang diusulkan Trivedi dan timnya dalam memperbaiki akurasi regresi adalah Spectral Clustering [1]. Pada penelitian ini, penulis mengembangkan perangkat lunak menggunakan metode yang diusulkan oleh Trivedi tersebut. Perangkat lunak yang dikembangkan digunakan untuk memprediksi nilai ujian akhir mahasiswa [1].

Artikel ini terdiri dari 6 topik bahasan. Bab 1 memaparkan tentang pendahuluan yang merupakan latar belakang dari penelitian. Metodologi yang digunakan akan dijelaskan pada bab 2. Bab 3 merupakan desain dari penelitian dalam prediksi nilai mahasiswa, sedangkan bab 4 merupakan penjelasan dari implementasi metodologi yang digunakan. Uji coba dan kesimpulan akan dipaparkan pada bab 5 dan bab 6, sedangkan saran dalam pengembangan penelitian dipaparkan pada bab 7.

II. METODOLOGI

A. Spectral Clustering

Clustering merupakan salah satu metode eksplorasi data yang digunakan dalam mencari pola yang ada pada suatu dataset. Pada umumnya pola tersebut dapat dilihat dari kesamaan sifat, karakteristik, atau ciri dari *record-record* pada dataset [2], [3].

Salah satu metode clustering adalah Spectral Clustering. Pada Spectral Clustering, pengelompokan didasarkan atas kesamaan antara setiap data. Kesamaan tersebut dilihat dari keterkaitan antara setiap data. Pada Spectral Clustering akan dibentuk sebuah graph dari data yang ada. Dimana verteks dari graph tersebut merupakan setiap record pada data. Edgennya berupa hubungan antar data yang biasanya bernilai jarak dari dua record yang berhubungan [4].

Langkah-langkah dalam melakukan Spectral Clustering adalah sebagai berikut [1], [5]:

1. Kontruksi graph similaritas dari dataset training. verteks pada *graph* tersebut merupakan representasi dari setiap *record* pada data training. Bobot dari tiap *edge* merupakan jarak antara satu verteks dengan verteks lainnya. Perhitungan jarak antar verteks menggunakan persamaan jarak exponential yang tertulis pada persamaan berikut [1]

$$w_{ij} = \exp\left(\frac{-\|s_i - s_j\|^2}{2\sigma^2}\right) \quad (1)$$

Setelah itu, bobot dari setiap edge yang ada dibentuk menjadi matriks weight. Dengan begitu matriks weight merupakan representasi graph similaritas dari dataset.

2. Dari matriks weight dihitung derajat dari setiap verteks dengan menjumlahkan bobot dari edge yang terhubung pada verteks yang bersangkutan. Dari derajat verteks tersebut dapat dibentuk matriks degree yang merupakan matriks diagonal yang berisi bobot setiap verteks.
3. Dibentuk normalisasi matriks Laplacian dengan menggunakan matriks weight (W) dan matriks degree (D) yang telah dihitung sebelumnya. Perhitungan matriks Laplacian (L) dengan rumus [6]

$$L_{sym} = D^{-1/2} W D^{-1/2} \quad (2)$$

4. Dihitung k eigen-vector pertama dari matriks Laplacian, dimana k merupakan parameter jumlah cluster. Maka terbentuklah matriks k-eigen yang merupakan k eigenvector pertama dari matriks Laplacian. Matriks k-eigen berukuran n x k, variabel n merupakan jumlah record pada data masukan.
5. Normalisasi data dengan matriks k-eigen sehingga akan terbentuk k kolom yang merepresentasikan setiap nilai normalisasi eigen pada setiap kolomnya.
6. Hasil dari data normalisasi kemudian di-cluster dengan K-Means Clustering. Data normalisasi mewakili masukan data latih. Data latih ke-i akan dimasukkan pada suatu cluster jika dan hanya jika data hasil normalisasi ke-i masuk pada cluster yang sama.

B. Bootstrap Aggregating Regresi Linier

Regresi Linier merupakan metode regresi dimana persamaan yang dihasilkan berupa persamaan linier. Dari persamaan yang dihasilkan dapat dihitung prediksi dengan memasukkan nilai-nilai variabel prediktor pada persamaan tersebut. Dari proses tersebut dapat dihasilkan nilai prediksi variabel respon. Multiple Regresi Linier dapat menggunakan lebih dari satu variabel prediktor dan satu variabel respon. Variabel prediktor dan respon yang digunakan pada Regresi Linier berupa variabel numerik.

Persamaan umum pada metode Multiple Regresi Linier adalah [7]

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_k x_k \quad (3)$$

Algoritma dalam mengestimasi koefisien Multiple Regresi Linier adalah least square estimator (LSE).

Metode *Bootstrap Aggregating* yang diterapkan pada Regresi Linier digunakan untuk memperbaiki performa model prediksi. Data yang akan diproses dibagi menjadi beberapa bagian kemudian dibentuk persamaan regresi dari bagian-bagian tersebut, sehingga akan dihasilkan persamaan regresi sebanyak jumlah bagian dari data tersebut. Saat ada data yang akan diprediksi, maka data dihitung dengan keseluruhan persamaan regresi dan hasilnya dirata-rata sehingga menjadi satu nilai yang merupakan nilai prediksi pada data tersebut [8].

III. PREDIKSI NILAI MAHASISWA

Prediksi nilai mahasiswa dilakukan dengan dua tahap yaitu tahap latih dan tahap uji. Tahap latih merupakan tahap pembelajaran yang dilakukan oleh sistem. Pada tahap ini dibentuk model prediksi dari data latih yang akan digunakan pada tahap uji. Masukan pada tahap latih berupa data latih dengan parameter jumlah cluster (k). Keluaran dari tahap ini adalah persamaan regresi sebanyak k. Tahap latih ini dibagi menjadi dua tahap, yaitu tahap Clustering dengan Spectral Clustering dan tahap regresi dengan Bootstrap Aggregating Regresi Linier.

Dari data masukan yang ada berupa nilai mahasiswa maka di-cluster menjadi beberapa kelompok sesuai dengan parameter jumlah cluster (k) yang dimasukkan. Metode clustering yang digunakan adalah Spectral Clustering. Dari clustering tersebut maka akan dihasilkan k-cluster data nilai mahasiswa. Pada masing-masing cluster-nya kemudian dimodelkan dengan persamaan Multiple Regresi Linier. Regresi menggunakan variabel prediktor nilai evaluasi dan nilai akhir atau UAS sebagai variabel respon. Dari persamaan-persamaan regresi tersebut maka dapat digunakan dalam memprediksi nilai mahasiswa dengan prediktor yang sama yaitu nilai evaluasi. Untuk mendapatkan satu nilai prediksi pada satu record maka digunakan metode Bootstrap Aggregating pada Regresi Linier.

Tahap uji dilakukan untuk memprediksi data uji menggunakan model yang telah dibuat pada tahap latih. Pada metode prediksi Bootstrap Aggregating Regresi Linier, variabel prediktor pada data uji yang berupa nilai evaluasi dimasukkan ke semua model hasil dari tahap latih. Jika ada k fungsi regresi yang dihasilkan pada tahap latih, maka setiap *record* pada data uji akan menghasilkan k prediksi nilai akhir.

Dari k prediksi nilai akhir dirata-rata untuk menghasilkan satu nilai sehingga didapatkan satu prediksi nilai akhir.

Selain itu, tahap uji berguna untuk menguji model yang dibuat pada tahap latih. Dari hasil prediksi yang ada dapat dihitung kesalahan hasil prediksi jika diketahui nilai sebenarnya. Nilai ujian akhir hasil prediksi dibandingkan dengan nilai ujian akhir sebenarnya. Metode yang digunakan adalah Root Mean Squared Error (RMSE) dengan persamaan [9]

$$RMSE = \frac{\sum_{i=1}^n \sqrt{(x_i - f_i)^2}}{n} \quad (4)$$

Semakin kecil nilai RMSE pada kesalahan hasil prediksi semakin kecil, begitu juga sebaliknya.

Pada tahap ini juga dilakukan metode Cross Validation dalam menguji model yang dihasilkan. Pada Cross Validation, dataset yang digunakan dibagi menjadi beberapa bagian (n). Pengujian dilakukan berulang dengan mengombinasikan data latih dan data uji pada dataset yang telah dibagi tersebut. Iterasi pengujian dilakukan sebanyak jumlah pembagian dataset (n). Data uji pada setiap iterasi ke-i merupakan bagian dataset ke-i dari pembagian dataset tersebut dan sisa dataset lainnya menjadi data latih.

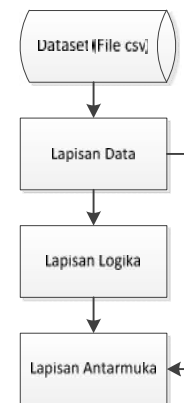
Data yang digunakan adalah kumpulan data yang berisi nilai mahasiswa. Nilai tersebut terdiri dari nilai-nilai evaluasi sebagai variabel prediktor dan nilai Ujian Akhir atau UAS sebagai kelas atribut yang akan diprediksi atau variabel respon. Dataset nilai mata kuliah yang digunakan ditunjukkan pada tabel 1.

Dari dataset tersebut terdapat dua jenis pengujian. Pengujian pertama dilakukan dengan menggunakan data masukan secara menyeluruh sebagai data latih dan data uji. Disamping itu, pada pengujian kedua, dataset2 dan dataset3 dibagi menjadi data latih dan data uji dengan metode Cross Validation.

Pada penelitian ini tidak dilakukan analisis terhadap missing value. Meskipun begitu masih ada kemungkinan terdapat missing value pada data. Missing value dapat terjadi saat mahasiswa tidak mengikuti evaluasi. Missing value yang terdapat pada data akan digantikan dengan nilai 0. Data yang bernilai 0 akan dihapus agar sistem tidak menghasilkan error yang besar. Record yang memiliki missing value lebih dari 50% dari variabel yang ada maka akan dihapus karena dianggap mahasiswa tersebut batal mengambil mata kuliah yang bersangkutan.

Tabel 1
Dataset Nilai Mata Kuliah

<i>Nama Dataset</i>	<i>Deskripsi</i>	<i>Jumlah Record</i>	<i>Variabel Prediktor</i>
dataset1	Nilai mata kuliah Statistika dan Komputasional tahun 2009/2010	40	4
dataset2	Nilai praktikum Struktur Data 2008/2009	180	5
dataset3	Nilai praktikum Struktur Data 2010/2011	153	3



Gambar 1 Struktur umum implementasi perangkat lunak prediksi nilai mahasiswa

IV. IMPLEMENTASI

Dalam implementasi metode dibuat perangkat lunak yang dibagi menjadi tiga lapisan yaitu lapisan data, logika serta lapisan antarmuka. Struktur umum dari aplikasi ini dapat dilihat pada gambar 1.

Lapisan data menangani pengolahan data serta struktur data pada perangkat lunak. Lapisan data ini memberikan layanan ke lapisan lainnya. Lapisan data terhubung dengan entitas dataset. Dataset yang digunakan adalah daftar nilai mahasiswa yang disimpan dalam *file csv*.

Lapisan logika menangani algoritma yang digunakan yaitu algoritma *Spectral Clustering* dan *Bootstrap Aggregating Regresi Linier*. Fungsi utama *Spectral Clustering* diimplementasikan pada fungsi *cluster* yang ditunjukkan pada gambar 2. Fungsi *cluster* memanggil fungsi *getEigen* untuk mendapatkan k (jumlah *cluster*) nilai eigen pertama dari matrix laplacian yang telah dikonstruksi. Fungsi *cluster* menggunakan algoritma *K-Means Clustering* dalam mengelompokkan.

Fungsi estimasi koefisien regresi pada algoritma *Bootstrap Aggregating Regresi Linier* diimplementasikan pada fungsi *regress* yang dapat dilihat pada gambar 3.

Lapisan terakhir adalah lapisan antarmuka yang menangani antarmuka antara aplikasi dengan pengguna. Lapisan antarmuka merupakan jembatan antara pengguna dengan perangkat lunak.

```

public void cluster() {
    getEigen();
    normalizeRow();
    KMeansClustering cluster = new
    KMeansClustering(numCluster, firstEigArr);
    cluster.cluster();
    this.setMember(cluster.getMember());
    setListMember(cluster.getListMember());
}
  
```

Gambar 2 Implementasi metode Spectral Clustering pada fungsi cluster

```

public double[] regress(DataValue dv) {
    int row = dv.size();
    int col = dv.get(0).size();
    double[][] x = new double[col][col];
    double[][] y = new double[col][1];
    for (int i = 0; i < col; i++) {
        for (int j = 0; j < col + 1; j++) {
            double sum = 0;
            if (i == 0 && j == 0) {
                sum = row;
            } else if (j == 0) {
                for (int k = 0; k < row; k++)
                    sum += dv.get(k).get(i-1);
            } else if (i == 0) {
                for (int k = 0; k < row; k++)
                    sum += dv.get(k).get(j-1);
            } else {
                for (int k = 0; k < row; k++)
                    sum += dv.get(k).get(i-1)
                        * dv.get(k).get(j-1);
            }
            if (j == col) x[i][1] = sum;
            else y[i][j] = sum;
        }
    }
    Matrix A = new Matrix(x);
    Matrix B = new Matrix(y);
    Matrix R = A.solve(B);
    double val[] = new double[col];
    for (int i = 0; i < col; i++)
    {
        val[i] = R.get(i, 1);
    }
}

```

Gambar 3 Implementasi metode Regresi Linier pada fungsi regress

V. UJICOB

Uji coba dilakukan dengan melatih data menggunakan parameter jumlah cluster yang berbeda-beda. Parameter jumlah cluster yang digunakan pada uji coba adalah 1, 2, 3, 4, 5, dan 10 cluster.

Terdapat dua jenis pengujian. Pada pengujian pertama data masukan digunakan sebagai data latih dan data uji sekaligus. Pengujian kedua menggunakan Cross Validation pada dataset2 dan dataset3. Jumlah fold yang digunakan pada Cross Validation adalah 5. Dataset2 dan dataset3 akan dibagi menjadi 5 bagian data yang akan bergantian menjadi data uji. Dari setiap parameter jumlah cluster akan dihasilkan 5 uji coba dengan menggunakan Cross Validation. Pada setiap pengujian akan dilakukan uji coba dengan menggunakan dua metode Clustering yaitu Spectral Clustering dan K-Means Clustering.

Untuk setiap pasangan data masukan dan data keluaran pada setiap parameter jumlah cluster akan dihitung kesalahan hasil prediksi menggunakan Root Mean Square Error (RMSE). Pada setiap dataset dihitung rata-rata RMSE. Hasil perbandingan nilai RMSE pada uji coba I dapat dilihat pada Tabel 2. Pada pengujian Cross Validation akan dihitung rata-rata RMSE untuk setiap parameter jumlah cluster. Tabel 3 menunjukkan rata-rata dari setiap parameter jumlah cluster yang digunakan pada uji coba dua.

Tabel 2
Hasil uji coba I pada dataset1, dataset2, dan dataset3

Jumlah Cluster (k)	dataset1		dataset2		dataset3	
	Spectral	K-Means	Spectral	K-Means	Spectral	K-Means
1	1.51E-16	1.51E-16	0.048	0.048	0.070	0.070
2	1.35E-16	9.47E-17	0.050	0.314	0.073	0.110
3	1.89E-16	2.05E-02	0.082	0.170	0.071	0.087
4	1.26E-16	1.85E-02	0.050	0.110	0.071	0.098
5	1.21E-02	1.48E-02	0.050	0.090	2.185	0.112
10	0.037	0.068	0.053	0.088	0.073	0.105
Rata-rata RMSE	8.28E-03	2.04E-02	0.055	0.137	0.424	0.097

Dari uji coba yang telah dilakukan, hubungan antara jumlah cluster terhadap hasil prediksi tidak dapat ditentukan secara pasti. Pengaruh jumlah cluster pada hasil prediksi bergantung pada masing-masing dataset yang digunakan. Tetapi pada umumnya semakin besar jumlah cluster maka semakin besar kesalahan hasil prediksi pada fase-fase tertentu. Misalnya uji coba 2 metode Spectral Clustering pada dataset3, nilai RMSE meningkat pada jumlah cluster 1 dan 2. Kemudian nilai RMSE turun pada cluster 3 dan meningkat lagi pada cluster 4. Tetapi tidak menutup kemungkinan semakin besar jumlah kluter menyebabkan nilai kesalahan semakin kecil, seperti pada uji coba I dataset1. Hal ini tidak menutup kemungkinan juga pada suatu kasus tidak ada pola yang terbentuk antara jumlah cluster dan nilai kesalahan hasil prediksi.

Pada beberapa kasus terdapat kesalahan hasil prediksi yang sangat tinggi dengan parameter jumlah cluster tertentu. Misal pada uji coba I dataset3, nilai RMSE pada parameter jumlah cluster 5 menghasilkan kesalahan yang sangat tinggi dan berbeda jauh dari kesalahan hasil prediksi yang dihasilkan dengan jumlah cluster lainnya. Hal ini dikarenakan pada metode Bootstrap Aggregating Regresi Linier nilai prediksi bergantung pada semua model regresi yang ada. Model prediksi pada uji coba tersebut menghasilkan nilai prediksi yang beragam dengan interval yang jauh. Sebagai contoh pada record pertama dengan nilai prediktor 74.9, 76 dan 77 serta variabel respon 80.5. Dari data tersebut, dari cluster 1 sampai cluster 4 menghasilkan nilai prediksi 78.46, 77.48, 76.86, 76.35, dan -199.7. Nilai prediksi yang dihasilkan 21.89 yang merupakan rata-rata dari hasil masing-masing cluster.

Karakteristik dari masing-masing persamaan regresi pada satu model bergantung pada persebaran data dari masing-masing cluster-nya, sehingga pada satu data dapat dihasilkan nilai prediksi yang cukup bagus pada satu persamaan tetapi juga dapat dihasilkan nilai prediksi yang jauh dari harapan bahkan melebihi interval dari suatu nilai (0-100).

Tabel 3
Hasil uji coba II (Cross Validation) pada dataset2 dan dataset3

Jumlah Cluster (k)	dataset2		dataset3	
	<i>Spectral</i>	<i>K-Means</i>	<i>Spectral</i>	<i>K-Means</i>
1	0.041	0.049	0.059	0.081
2	0.047	0.206	0.060	0.088
3	0.052	0.120	0.065	0.124
4	0.053	0.109	0.060	0.162
5	0.058	0.139	0.078	0.211
10	0.075	0.162	0.081	0.153
Rata-rata RMSE	0.054	0.131	0.067	0.137

Secara umum hasil prediksi yang dihasilkan metode *Bootstrap Aggregating Regresi Linier* dengan *Spectral Clustering* lebih baik daripada *K-Means Clustering*. Hal ini disebabkan karena pengelompokan pada metode *Spectral Clustering* memperhatikan hubungan antara masing-masing data pada dataset yang digunakan.

Disamping itu pada dunia nyata, persebaran data sangatlah bervariasi, termasuk dengan data nilai mahasiswa. Pada *K-Means Clustering* pengelompokan dilakukan berdasarkan jarak terdekat dari suatu *cluster* (*centroid*). Pada *Spectral Clustering* pengelompokan dilakukan dengan mempertimbangkan jarak atau hubungan antar masing-masing data.

Persebaran data dari suatu dataset mempengaruhi hasil prediksi yang dihasilkan. Dari uji coba diatas, kesalahan pada dataset1 jauh lebih kecil dibandingkan dengan dataset2 dan dataset3. Hal ini bisa terjadi karena persebaran data pada dataset1 lebih merata dibandingkan lainnya. Jumlah record juga berperan dalam perhitungan kesalahan hasil prediksi. Jumlah record pada dataset1 lebih sedikit dibandingkan dengan dataset2 serta dataset3 sehingga dimungkinkan variasi data yang ada pada dataset1 lebih sedikit daripada dataset2 atau dataset3.

Hal lain yang berpengaruh pada hasil prediksi adalah jumlah atribut prediktor yang terlibat. Dari dataset2 dan dataset3 yang memiliki jumlah record yang banyak, variabel prediktor pada dataset2 lebih banyak daripada dataset3. Dari uji coba secara umum nilai RMSE pada dataset2 lebih kecil dibandingkan uji coba pada dataset3. Hal ini disebabkan karena dengan banyaknya variabel prediktor maka pertimbangan hasil prediksi juga semakin banyak. Selain itu, persamaan regresi yang dihasilkan lebih dinamis dengan variabel prediktor yang lebih banyak.

VI. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa kesimpulan yang dapat diambil, yaitu:

1. Perangkat lunak yang dikembangkan dengan algoritma *Spectral Clustering* yang mendukung algoritma *Bootstrap Aggregating Regresi Linier* terbukti mampu melakukan

prediksi nilai mahasiswa. Hal ini terlihat dari nilai kesalahan RMSE sekitar 0.05 – 0.08 dari dataset yang digunakan.

2. Pada uji coba yang dilakukan, perangkat lunak yang memanfaatkan algoritma *Spectral Clustering* yang mendukung *Bootstrap Aggregating Regresi Linier* memiliki performa yang lebih baik jika dibandingkan dengan perangkat lunak yang menggunakan algoritma *K-Means Clustering*.
3. Parameter jumlah cluster yang tidak tepat dapat menyebabkan kesalahan hasil prediksi yang cukup tinggi.
4. Dari uji coba yang dilakukan, jumlah atribut prediktor yang lebih banyak dapat menghasilkan hasil prediksi menjadi lebih baik.

VII. PENGEMBANGAN PENELITIAN

Prediksi nilai akademik mahasiswa dapat dikembangkan sehingga menghasilkan nilai prediksi yang lebih baik. Salah satu pengembangan dapat dilakukan dengan algoritma penentuan jumlah *cluster* karena parameter jumlah *cluster* berpengaruh terhadap hasil prediksi. Disamping itu dapat dilakukan pengembangan dalam *Bootstrap Aggregating Regresi Linier* dengan dilakukan pemilihan terhadap model regresi yang digunakan. Hal ini dikarenakan model regresi yang dihasilkan belum tentu sesuai dengan data yang akan diuji.

DAFTAR PUSTAKA

- [1] Trivedi, S., A. Pardos, Z., & N. Sar, G, *Spectral Clustering in Educational Data Mining*, (2008).
- [2] Tan, P.-N., Steinbach, M., & Kumar, V., *Introduction to Data Mining*. Minnesota: Addison-Wesley, (2006).
- [3] Zhang, Z., Wu, X., & S. Yu, P., *Spectral Clustering for Multi-type Relational Data*, (2006).
- [4] Y. Ng., A., I. Jordan, M., & Weiss, Y., *On Spectral Clustering : Analysis and Algorithm*, (2002).
- [5] von Luxburg, U., *A Tutorial on Spectral Clustering*. 17, (2007) 4.
- [6] Mohar, B., *The Laplacian Spectrum of Graphs*, In *Graph Theory*, (1991) 871-898.
- [7] E. Walpole, R., H. Myers, R., & L. Myers, S., *Probability & Statistics for Engineers & Scientists*. Pearson Education International, (2007).
- [8] Chen, T., & Ren, J., *Bagging for Gaussian process regression*, (2008).
- [9] Karagozoglu, B., & Turkmen, N., *A software tool to facilitate design, assessment and evaluation of courses in an educational system*, (2007).